National Leadership Computing Facility - *Bringing Capability Computing to Science*

Douglas B. Kothe

Director of Science

National Center for Computational Sciences

Oak Ridge National Laboratory

ORNL is DOE's largest multipurpose science laboratory

- $1.06 billion budget
- 3,900 employees
- 3,000 research guests annually
- Nation's largest open scientific computing facility
- Nation's largest science facility: the $1.4 billion Spallation Neutron Source
- Nation's largest concentration of open source materials research
- Nation's largest energy laboratory
- $300 million modernization in progress

Computing and Computational Sciences Directorate

Associate Laboratory Director
Thomas Zacharia

Linda Malone, Executive Secretary

National Center for Computational Sciences
Jeff Nichols, Acting Director

Computational Sciences and Engineering Division
Brian Worley, Director

Computer Science and Mathematics Division
Jeff Nichols, Director

Networking and Computing Technologies Division
Becky Verastegui, Director

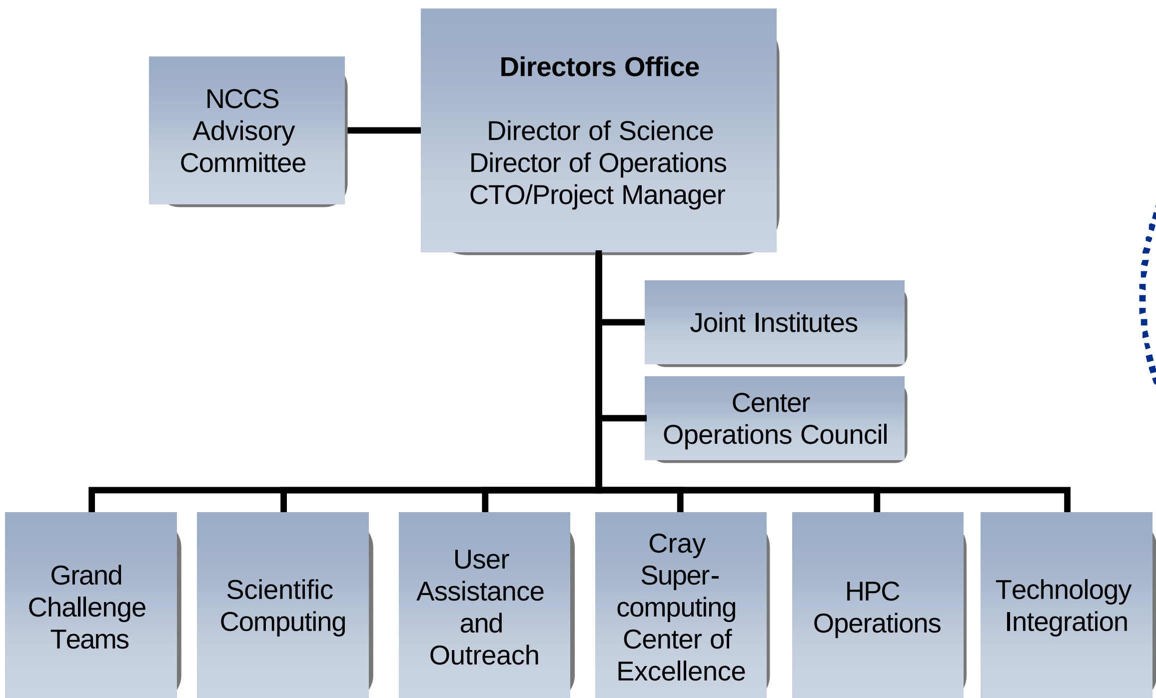
Matrix Support
William Besancenez
Contracts
Dave Green
Recruiting
Janet Lu, Kim Milburn, Matt Rogers
Finance Managers
Dave Rupert
HR Manager
Ronnie Wagoner
Business Manager
Jack Wells
CCS Fellow

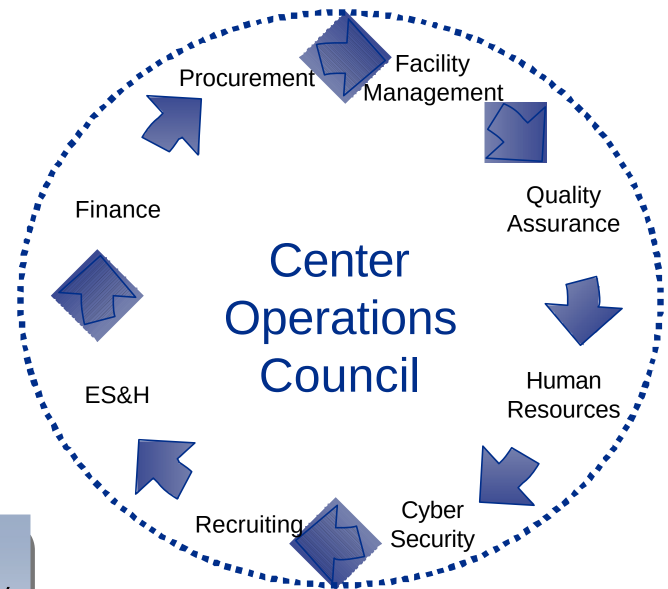
Staff
Debbie McCoy
Administration and Coordination
Human Resources
Nancy Wright
Human Resources

Advanced Scientific Computing Research Program
Thomas Zacharia

NCCS Organization

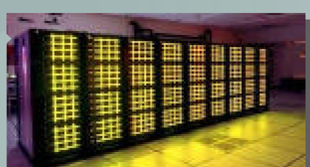
An organizational focused engaging the scientific and engineering communities to provide a new computational capability designed to maximize the performance of scientific applications.





National Center for Computational Sciences performs three inter-related activities for DOE

- Deliver National Leadership Computing Facility for science
  - **Focused on grand challenge science and engineering applications**
- Principal resource for SciDAC and (more recently) other SC programs
  - **Specialized services to the scientific community: biology, climate, nanoscale science, fusion**
- Evaluate new hardware for science
  - **Develop/evaluate emerging and unproven systems and experimental computers**



Intel Paragon: World's fastest computer

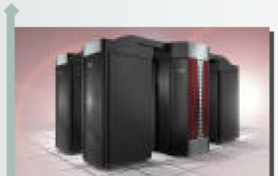
1995

IBM Power3: DOE-SC's first terascale system

2000

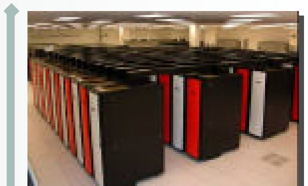
IBM Power4: 8th in the world (2001)

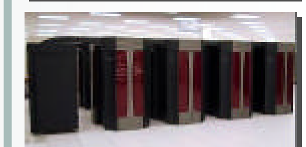
2001



Cray X1: Capability computer for science

2004





Cray XT3 and X1E: Leadership computers for science

2005

Integrate core capabilities to deliver computing for frontiers of science

Develop and evaluate next-generation architectures with industry

IBM

CRAY

sgi

Provide leadership-class computing resources for the Nation

Create math and CS methods to enable use of resources

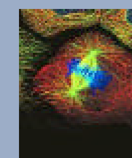
SciDAC ISICs

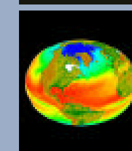
Scientific Applications Partnerships

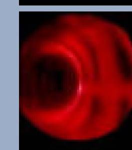
Modeling and simulation expertise

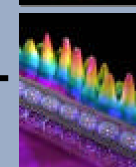
Transform scientific discovery through advanced computing

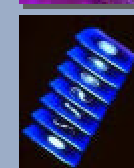
Computational End Stations


Biology


Climate


Fusion


Materials

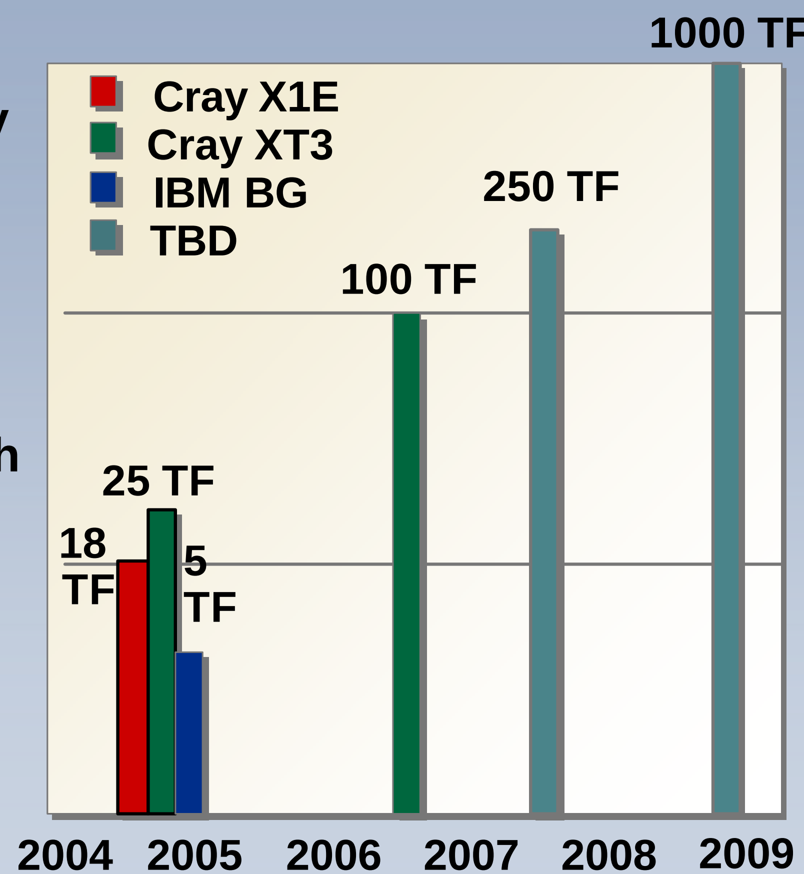

Industry/ other agencies

NLCF plan for the next 5 years:

Vector Arch
Global memory
Powerful CPU

Cluster Arch
Low latency
High bandwidth

Scalability
100K CPU
MB/CPU



New world-class facility capable of housing leadership class computers

- $72M private sector investment in support of leadership computing
- Space and power:
  - **40,000 ft^2 computer center with 36-in. raised floor, 18 ft. deck-to-deck**
  - **8 MW of power (expandable) @ 5c/kWhr**
- High-ceiling area for visualization lab (Cave, Powerwall, Access Grid, etc.)
- Separate lab areas for computer science and network research

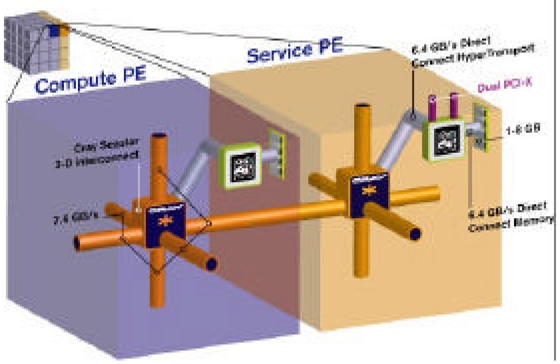
NCCS Cray X1E – Phoenix

- Largest Cray X1E in the world – 18.5TF
- 1024 processors – 400 MHz, 800 MHz vector units
- 2 TB globally addressable memory
- 32 TB of disk
- Most powerful processing node
  - 12.8 GF CPU, 2-5x commodity processors
- Highest bandwidth communication with main memory
  - 34.1 GB/sec

Highly scalable hardware and software
High sustained performance on real applications

NCCS Cray XT3 – Jaguar

Cray XT3 Scalable Architecture

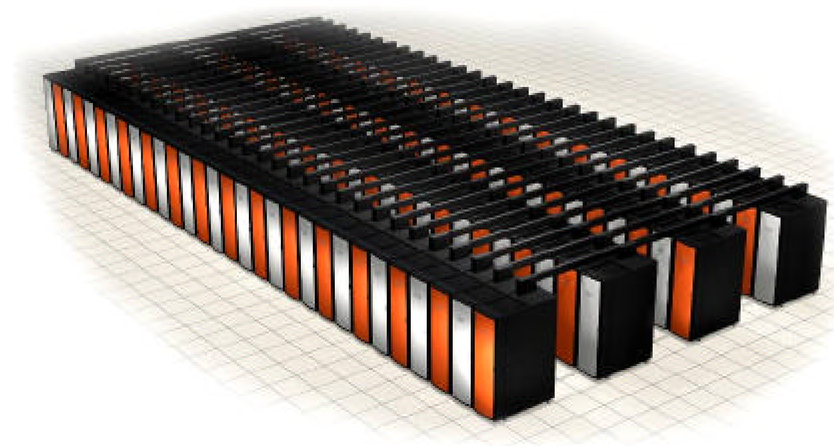


Accepted in 2005 and routinely running applications requiring 4,000 to 5,000 processors.

System Statistics	
Cabinets	56
Compute Processors	5,212 2.4 GHz Opteron
Lustre Object Storage Servers	58
10 Gigabit Ethernet nodes	2
System Services Nodes	8
Disk space	120 TB
Power	900 Kilowatts
Peak Performance	25.1 TeraFLOP/s

Future development of Jaguar

ORNL intends to expand Jaguar to a 100 Teraflop system in 2006 by doubling the number of cabinets and going to dual-core processors.



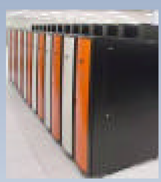
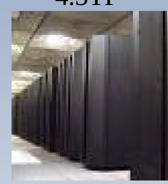
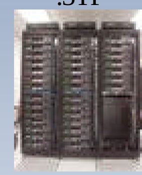
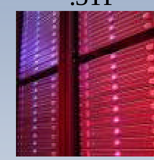
Pending approval by the U.S. Department of Energy and appropriation of the money by the Congress.

Cabinets	120
Compute Processors	Approximately 22,456
Memory	Approx. 45 TB (2 GB per processor)
Disk	480 TB
Peak Performance	100+ TeraFLOP/s

Where we are today

Network Routers

Control Network
1 GigE
10 GigE
UltraScience

Cray XT3 Jaguar	Cray X1E Phoenix	SGI Altix Ram	IBM SP4 Cheetah	SGI Linux OIC	IBM Linux NSTG	Visualization Cluster	IBM HPSS
25TF	18TF	1.5TF	4.5TF	8TF	.3TF	.5TF	
							
5294 2.4GHz 11TB Memory	1024 .5GHz 2 TB Memory	(256) 1.5GHz 2TB Memory	(864) 1.3GHz 1.1TB Memory	(1376) 3.4GHz 2.6TB Memory	(56) 3GHz 76GB Memory	(128) 2.2GHz 128GB Memory	Many Storage Devices Supported
120TB	32TB	36TB	32TB	80 TB	4.5TB	9TB	5TB

Shared Disk

Scientific Visualization Lab
- 27 projector PowerWall

Test Systems
- 96 processor Cray XT3
- 32 processor Cray X1E*
- 16 Processor SGI Altix

Evaluation Platforms
- 144 processor Cray XD1 with FPGAs
- SRC Mapstation
- Clearspeed
- BlueGene (at ANL)

Backup Storage  5PB

September 2005 Summary

8 Systems

Supercomputers
8,998 CPUs
19TB Memory
58 TFlops

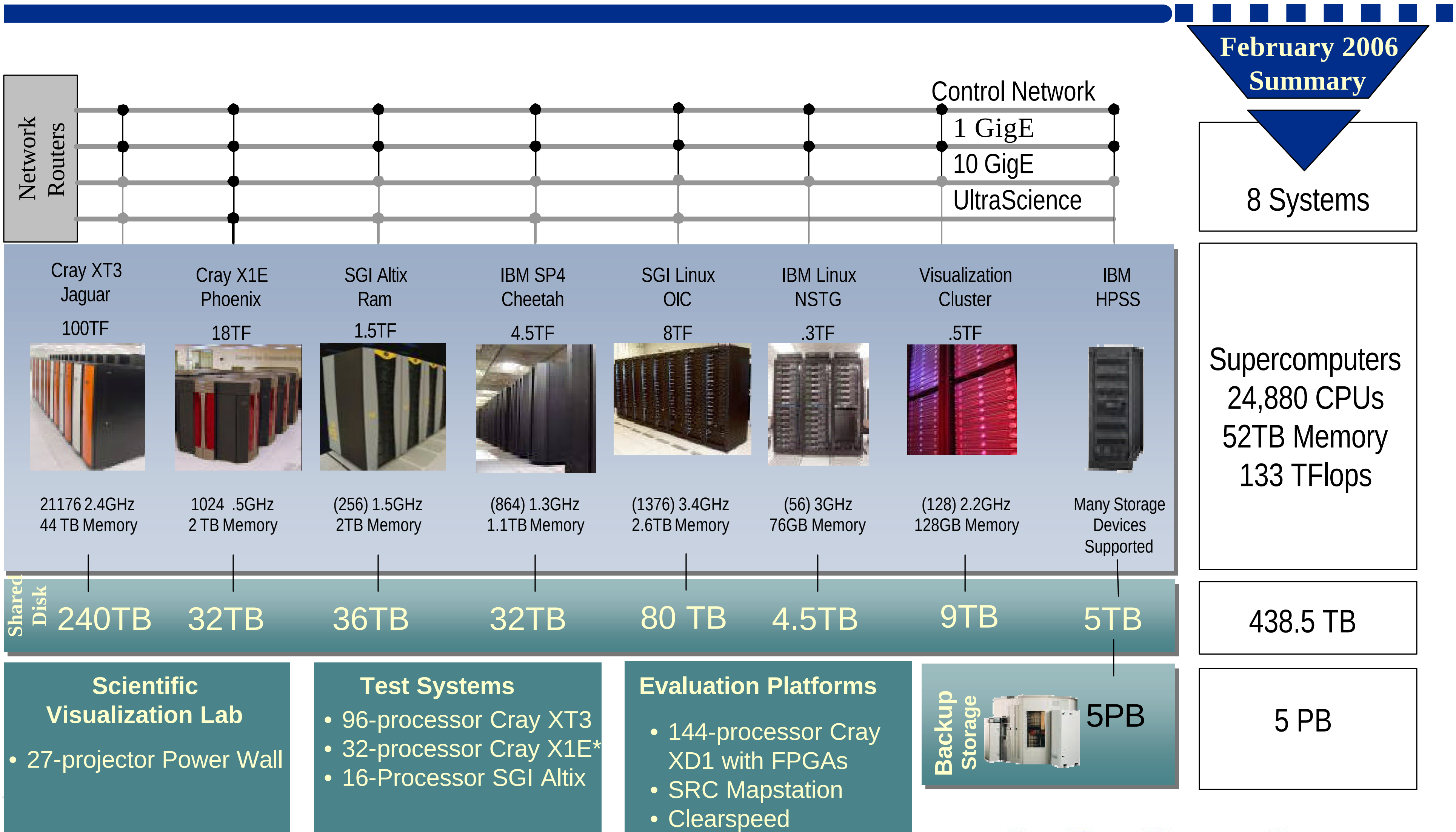
318.5 TB

5 PB

Computational Sciences

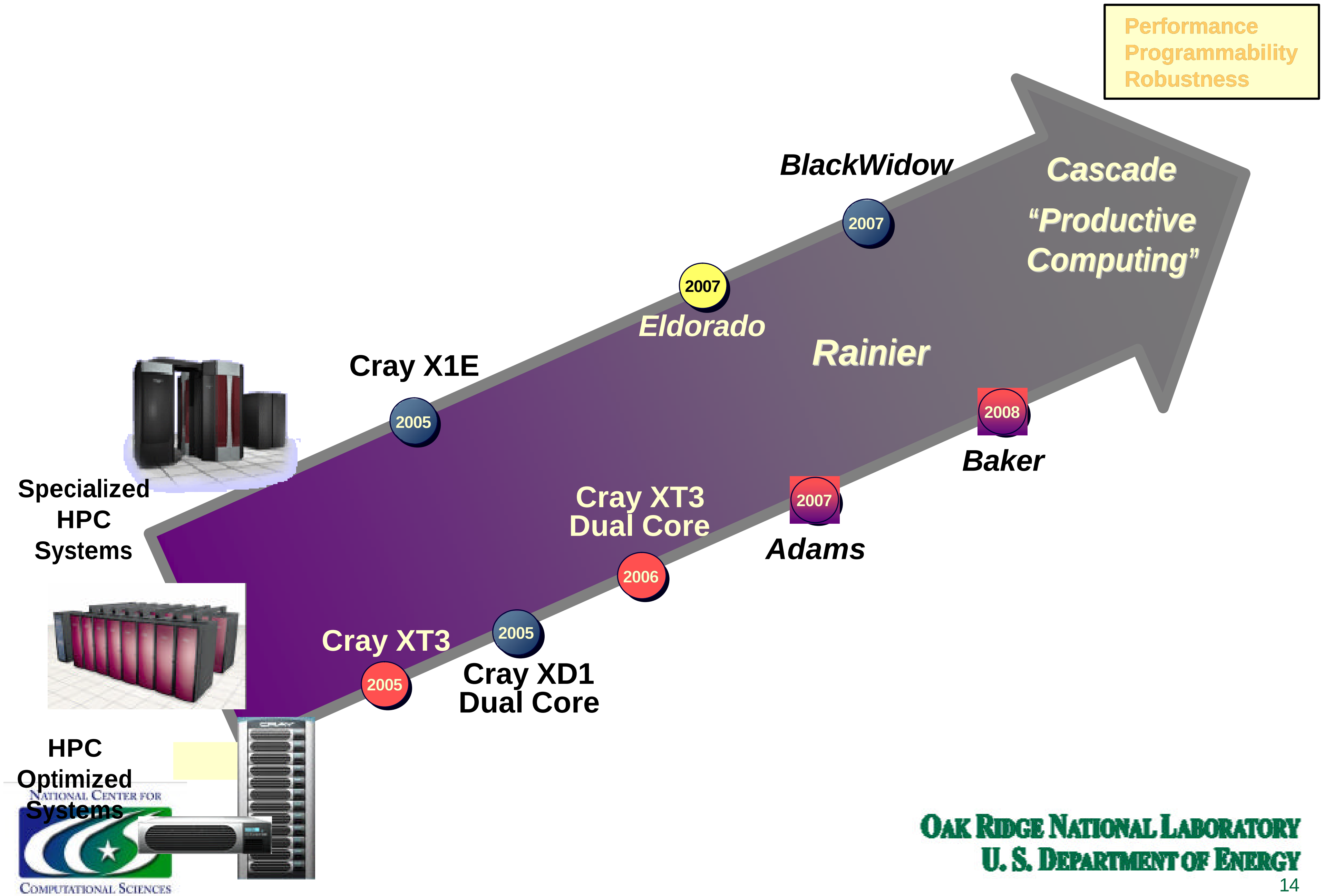
Oak Ridge National Laboratory
U. S. Department of Energy

Where we plan to be in 2006



ORNL Path to Leadership Computing

Scalable, High-Bandwidth Computing

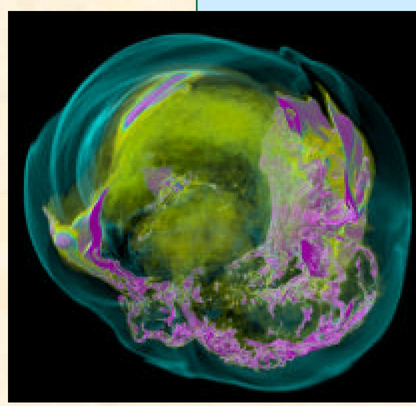


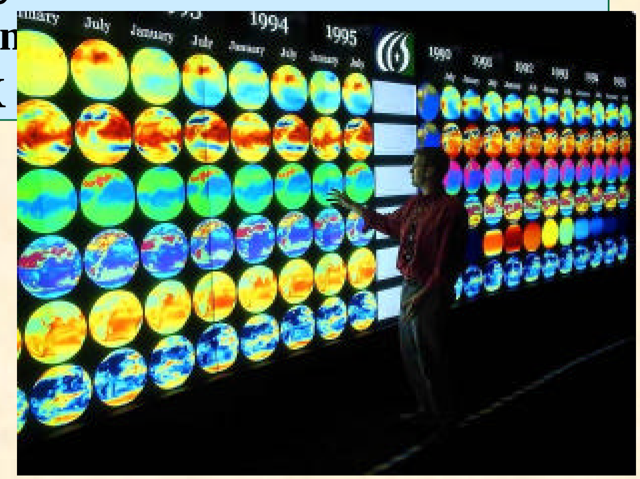
Visualization at the NCCS

- **“Production” visualization environment on visualization cluster**
  - **“Turnkey” visualization tools (VisIt, EnSight, Paraview, AVS, VMD, etc.)**
  - **Infrastructure integration (high-speed Quadrics interconnect, Lustre support, SLURM batch management)**
  - **Onsite support team (3.5 app developers, 1.5 researchers)**
  - **Large data emphasis**
- **Remote visualization capabilities**
  - **Client-server vis paradigm**
  - **Geometry-based, image**
  - **streaming**
- **Research into new vis techniques**
  - **Multiresolution streaming**
  - **Error-driven I/O**
  - **Statistical sci visualization**

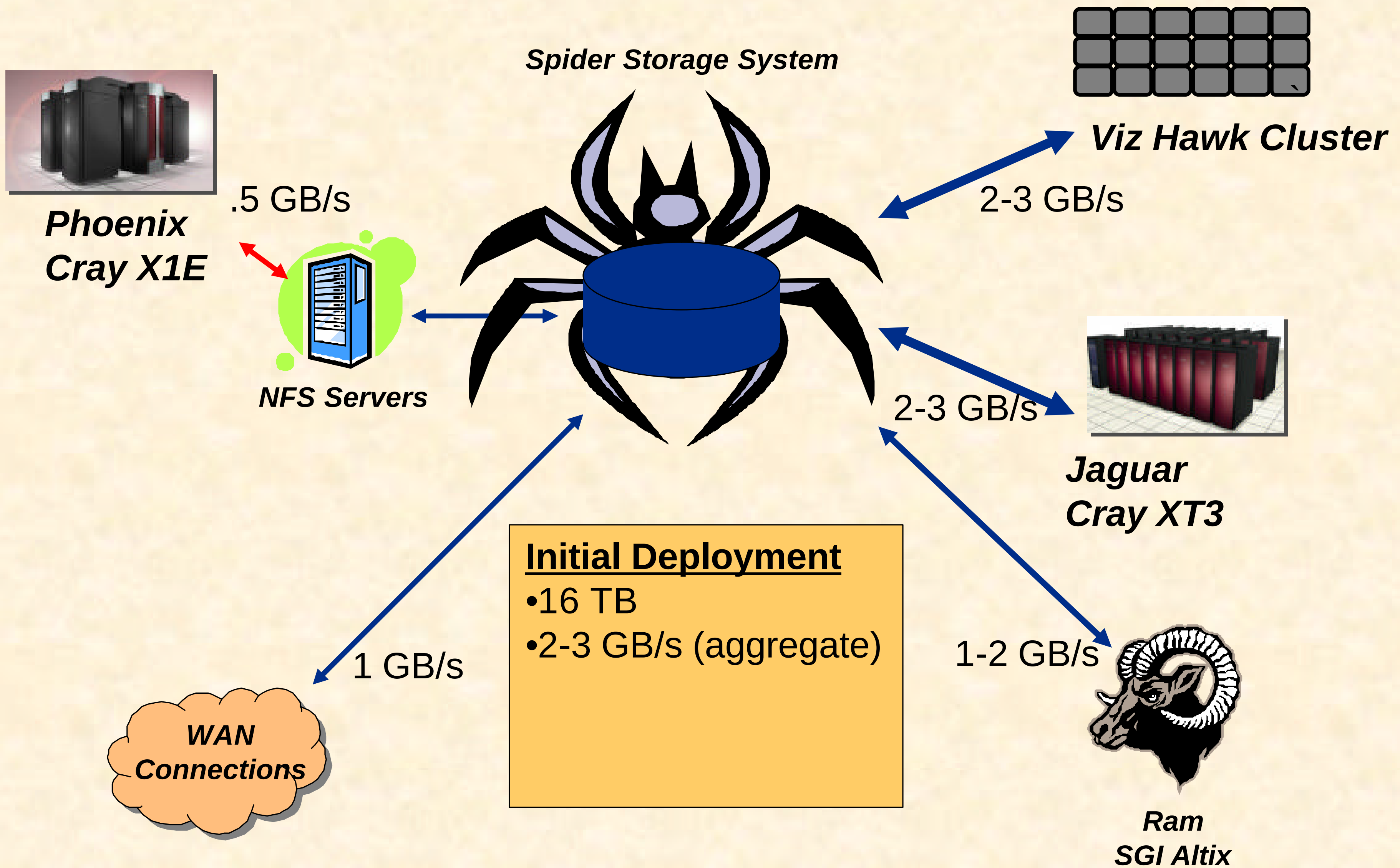
World-class display environment

- **EVEREST Powerwall**
  - **30’ x 8’ in size, 11k x 3k resolution, 35+ million pixels**
  - **Integrated into vis infratructure**
  - **Powerful discovery tool for research groups and collaborations**
- **Display middleware**
  - **Chron**
  - **MX**





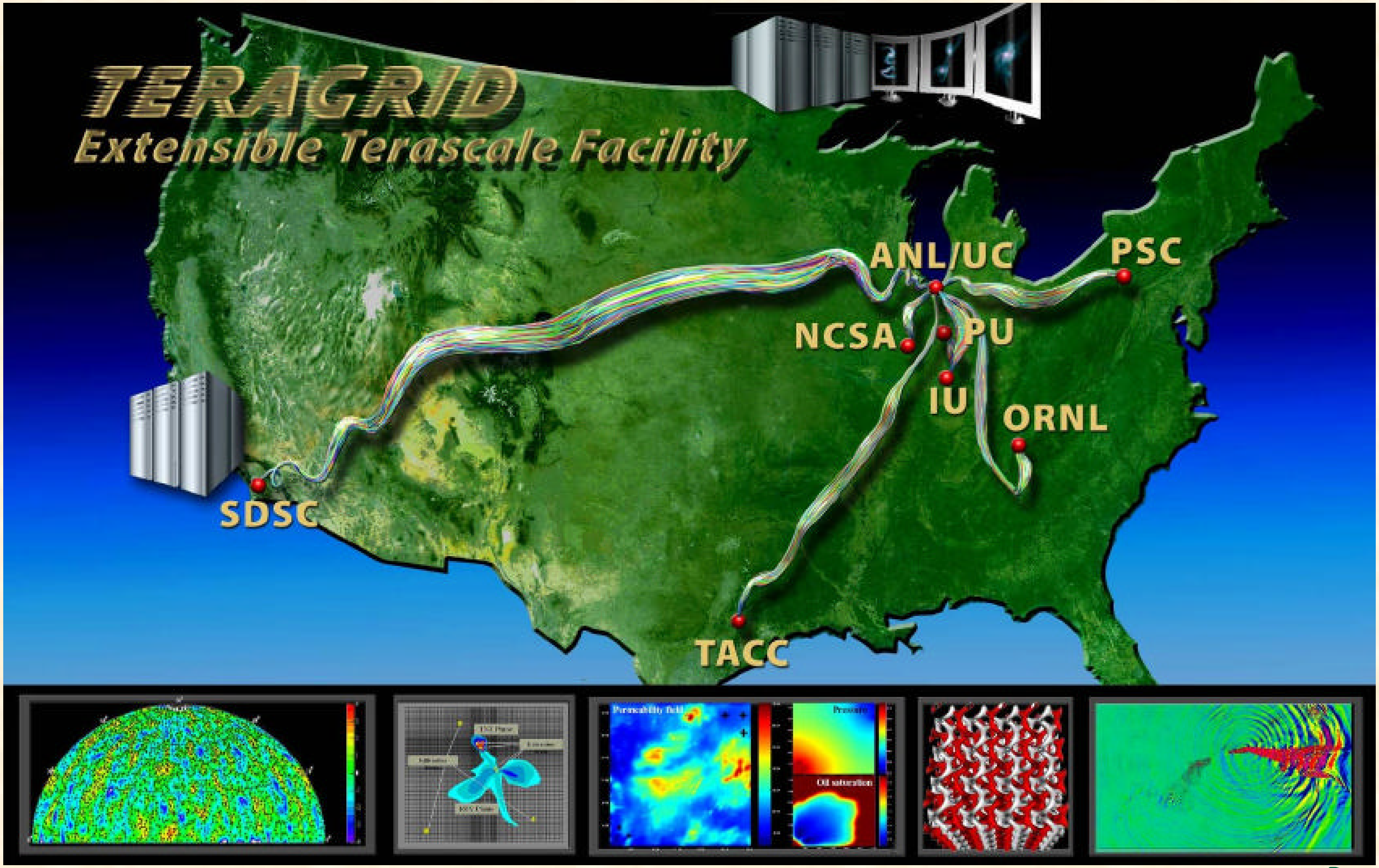
Spider - A Center Wide File System



<u>Initial Deployment</u>

- 16 TB
- 2-3 GB/s (aggregate)

Already a resource provider on TERAGRID



Information technology growth rate

Exponential growth creates the potential for revolutionary changes in what we do and how we do it

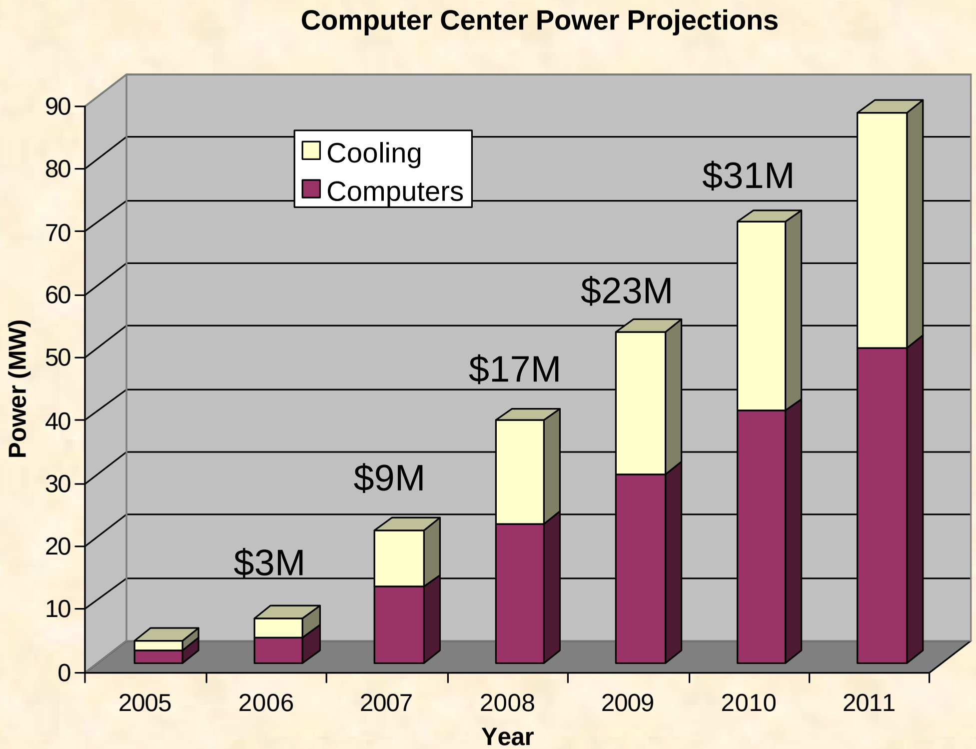
- **Processing power**
  - Doubling every 18 months
  - 60% improvement each year
  - Factor of 100 every decade
- **Disk Capacity**
  - Doubling every 12 months
  - 100% improvement each year
  - Factor of 1000 every decade
    - 10X as fast as processor performance!
- **Optical bandwidth today**
  - Doubling every 9 months
  - 150% improvement each year
  - Factor of 10,000 every decade
    - 10X as fast as disk capacity!
    - 100X as fast as processor performance!!

Computational science

Data science

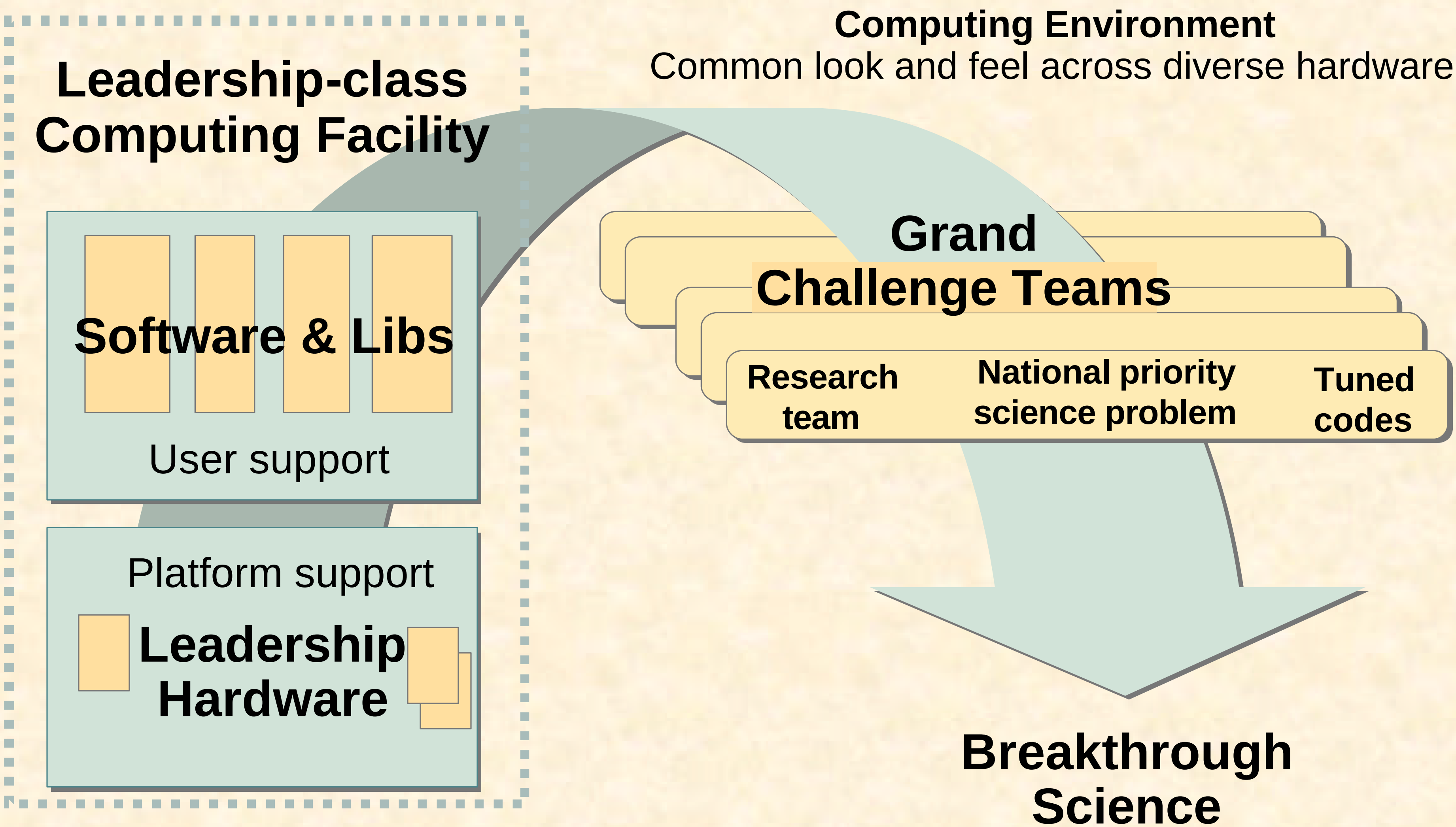
ORNL computing infrastructure needs Power and cooling 2006 - 2011

- **Immediate need to add 8 MW to prepare for 2007 installs of new systems**
- **NLCF petascale system could require an additional 10 MW by 2008**
- **Need total of 40-50 MW for projected systems by 2011**
- **Numbers just for computers: add 75% for cooling**
- **Cooling will require 12,000 – 15,000 tons of chiller capacity**



Cost estimates based on $0.05 kW/hr

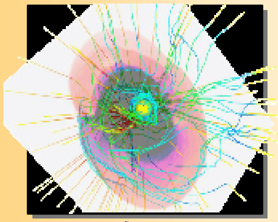
Facility plus hardware, software, science teams contribute to Science breakthroughs



Excerpts From ASCR FY07 President's Budget

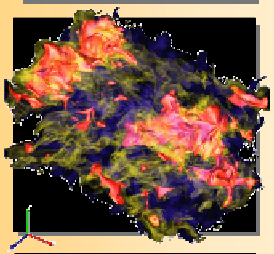
- "Computer-based simulation enables us to understand and predict the behavior of complex systems that are beyond the reach of our most powerful experimental probes or our most sophisticated theories."
- General Goal 5, World-Class Scientific Research Capacity
  - Provide world-class scientific research capacity needed to: ensure the success of DOE missions in national and energy security; advance the frontiers of knowledge in physical sciences and areas of biological, medical, environmental, and computational sciences; or provide world-class facilities for the Nation's science enterprise
- ASCR contributes to Goal 5 by enabling research programs across SC, as well as other elements of DOE, to succeed.
  - Develop multiscale mathematics, numerical algorithms, and software that enable effective models of systems such the earth's climate, the behavior of materials, or the behavior of living cells that involve the interaction of complex processes taking place on vastly different time and/or length scales
  - Develop, through the Genomics: GTL partnership with the Biological and Environmental Program, the computational capability to model a complete microbe and a simple microbial community.

FY05 X1E Allocations



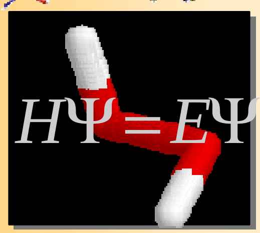
3D Studies of Stationary Accretion Shock Instabilities in Core Collapse Supernovae

A. Mezzacappa (Oak Ridge National Laboratory) and J. Blondin (North Carolina State University)
415,000 processor-hrs



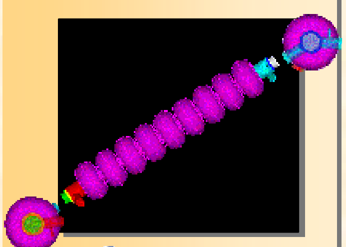
Turbulent Premix Combustion In Thin Reaction Zones

J.H. Chen (Sandia National Laboratories)
360,000 processor-hrs



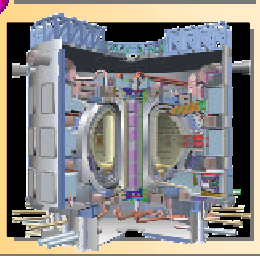
Full Configuration Interaction Benchmarks for Open Shell Systems

R. Harrison (Oak Ridge National Laboratory) and M. Gordon (Ames Laboratory)
220,000 processor-hrs



Computational Design of the Low-Loss Accelerating Cavity for the ILC

Kwok Ko (Stanford Linear Accelerator Center)
200,000 processor-hrs



Advanced Simulations of Plasma Microturbulence

W. M. Tang (Princeton University, Plasma Physics Laboratory)
50,000 processor-hrs

Cray XT3 Applications

Aero	GTC	Partisn	<u>Benchmarks</u>
Alegra	GYRO	POP	HALO
Amber/PMEMD	HYCOM	Presto	Hello World
AORSA	ITS	QCD-MILC	HPCC
ARPS	LAMMPS	Quake	HPL
AVUS (Cobalt-60)	Leonardo – Relativity Code	Quantum-ESPRESSO Suite	LINPACK
Calore	LM	S3D	NPB
CAM	LS-DYNA	Sage	OSU
CCSM	LSMS 1.6, 1.9, 2.0	Salinas	Pallas MPI
CHARM++	MAD9P	Siesta	PSTSWM
CHARMM	MILC	SPF	SMG2000
CPMD	moldyPSI	syr	sPPM
CTH	MPQC	TBLC	STREAM/triad
Dynamo	NAMD	Trilinos	Sweep3D
ECHAM5	NWChem	UMT2000	
FLASH	Overflow	VASP	
GAMESS	Paratec	WRF	
Gasoline – N-body astro.	Parmetis	ZEUS-MP	
Gromacs			9/1/05

55 applications + 13 Benchmarks

National Leadership Computing Facility
2006 Call for Proposals

- Projects must be funded by the DOE Office of Science or support the mission of the Office of Science
- Principal investigator, multi-principal investigator teams, multi-institutional teams and end station proposal teams may apply for LCF resources
- Multi-year proposals will be considered subject to annual review.

- Expectation that Leadership systems will enable U.S. to be "first to market" with important scientific and technological capabilities, ideas, and software
- Limited set of scientific applications selected and given substantial access to leadership system

Office of Science
U.S. DEPARTMENT OF ENERGY

Access to NLCF

Proposals

- Pilot Proje

NATIONAL CENTER FOR COMPUTATIONAL SCIENCES

Review

- Technical readin s
- Scalability

Allocations

- Grand Challenge
- End Stations
- Pilot Projects

Project Types

Grand Challenge

- Scientific problems that may only be addressed through access to LCF hardware, software, and science expertise
- Multi-year, multi-million CPU hour allocation

End Station

- Computationally intense research projects, also dedicated to development of community applications
- Multi-year, multi-million CPU hour allocation

Pilot Project

- Small allocations for projects, in preparation for future Grand Challenge or End Station submittals
- Limited in duration

Projects must

- possess quality of science (potential for "breakthrough")
- have a real need for leadership-class computer resources
- fulfill programmatic (DOE) or other Fed agency requirements

52 Responses to 2006 Call for Proposals

Life Sciences	2
Nanoscience	5
Materials	6
Computer	3
Chemical	5
Environmental	2
Engineering Physics	2
Computational Mechanics	1
Combustion	3
Climate and Carbon Research	6
Fusion	6
Astrophysics	6
Accelerator	2
Nuclear Physics	1
Turbulence	1
High Energy Physics	1

The majority of the proposals are led by DOE and University PIs.

FY06 LCF Project Portfolio by Science

- **Fusion (3)**
- **Climate Change (3)**
- **Astrophysics (3)**
- **Accelerator Physics (1)**
- **Computer Science (1)**
- **Nuclear Physics (1)**
- **Combustion (1)**
- **High Energy Physics (1)**
- **Materials Science (1)**
- **Life Sciences (1)**
- **Chemistry (1)**

FY06 INCITE Project Portfolio by Science

- **Fusion (1)**
- **Materials Science (1)**
- **Life Sciences (1)**
- **Computer Science (1)**
- **Engineering (1)**

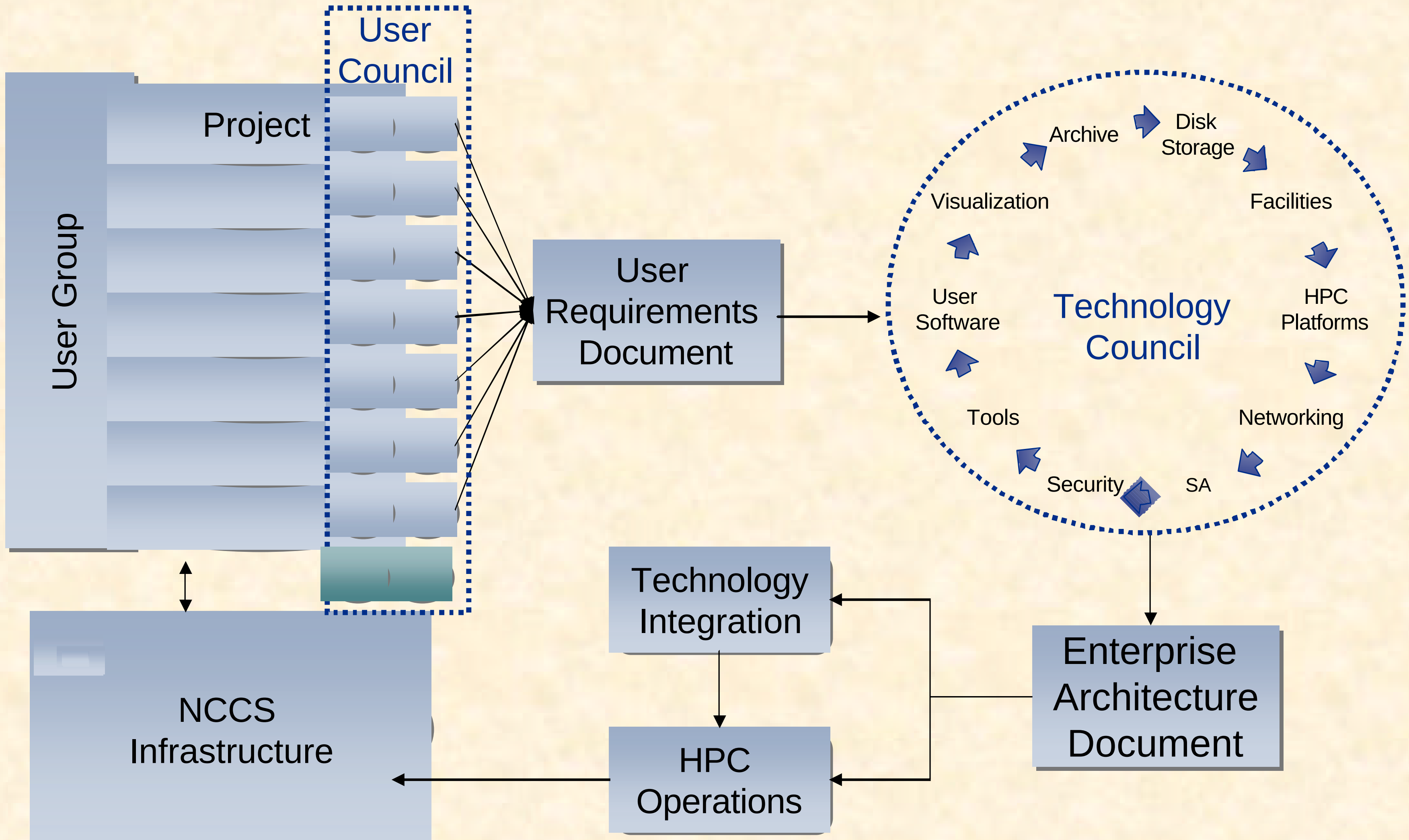
FY06 Project Portfolios

- **LCF Projects**
  - **Analysis: 6**
  - **Development and Analysis:11**
- **INCITE Projects**
  - **Analysis: 2**
  - **Development: 2**
  - **Development and Analysis: 1**
- **LCF Projects**
  - **Jaguar: 12 projects totaling 24,330,000 hours**
  - **Phoenix: 10 projects totaling 5,294,000 hours**
- **INCITE Projects**
  - **Jaguar: 3 projects totaling 2,935,000 hours**
  - **Phoenix: 2 project totaling 600,000 hours**

FY06 Project Portfolios: Observations

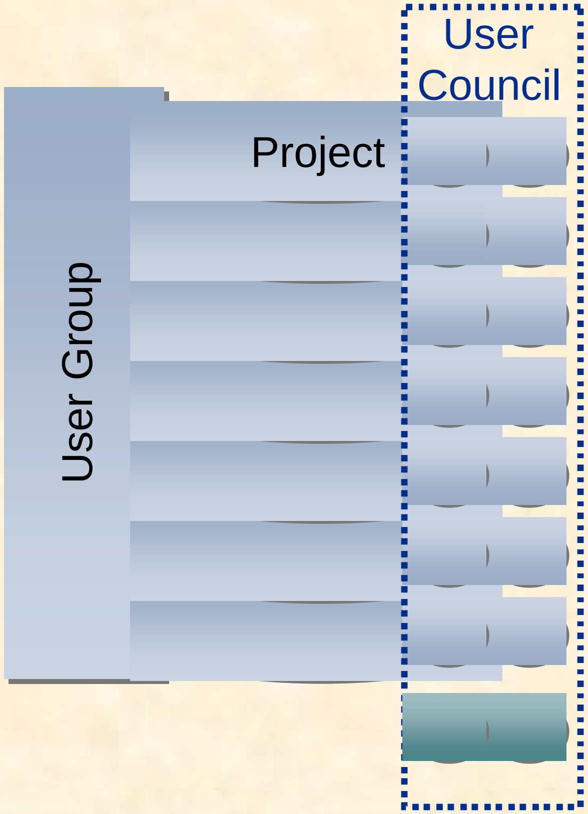
- **Need for Leadership-Class computers varies**
  - **New models, algorithms now made possible**
  - **Ability to get at unprecedented time and length scales (e.g., moving toward DNS)**
  - **Ability to explore more fully coupled physics**
  - **Faster (and now realistic) simulation turn-around time**
- **Many different types of metrics touted for increased "predictability" and understanding**
  - **But what does this mean? What are the requirements, consequences, and benefits?**
- **V&V**
  - **Is your V&V rigorous & quantitative enough? SQA adequate?**
  - **Is the V&V defensible and reproducible (testing)?**

NCCS – From User Requirements to an Enterprise Architecture



NCCS – User Council

Meets directly after allocations are awarded and frequently to adjust document.



An organizational model that will deliver:

- **Strong engagement of research community in setting research and operational requirements**
- **Direct points of contact into each NLCF grand challenge project.**
- **Focused on "future" requirements**
- **Two additional members appointed by NCCS Advisory and NCCS Director to cover emerging application areas**

Each NLCF Grand Challenge project will have one member on the User Council along with that projects liaison from the NLCF Scientific Computing team. These two individuals will ensure the user requirements document is up to date with the requirements from that project.

User Council – Feb 06

- **Membership**
  - **One rep per LCF/INCITE project team and its associated NCCS liaison**
  - **Co-chaired by NCCS (Doug Kothe) & User Group Chair (Jerry Bernholc)**
- **Communication**
  - **Monthly telecon? Individual phone calls, emails, face-to-face mtgs**
  - **This will not be as free form as the Users Group (focus on future)**
  - **Co-chairs will set agenda**
- **Product**
  - **Communication and advocacy of our mutual goals and needs**
  - **Annual NCCS requirements doc ("Green Book")**
  - **Influence next generation systems before delivery**

User Council Charter

- **Apps-driven requirements for future systems**
  - Constrained (driven by what will be coming) or
  - Unconstrained (what you would like to have achieve science breakthrough)
  - Try to look at the “hows” in addition to the “whats”
  - Start by gathering metrics, diving down into current state
- **Track emerging developments in your field**
- **Participate in and/or influence advanced architecture evaluation & design**
- **Be aware of emerging fields ripe for large-scale computational science that can benefit from our efforts**

Future Requirements

- **Consider all aspects of the end-to-end solution: before, after, & underneath the App**
  - **Hardware**
    - **“I need about 100 kB/unknown of on-node memory ”**
    - **“I need 0.01 μs/cell/cycle on-node performance”**
    - **“I need 1000 nearest-neighbor gathers per every unknown for every integration time step”**
  - **Software (everything below the App)**
    - **“The Linux kernel must support dynamic libraries”**
    - **“I need OpenMPI”**
  - **Methods/Models**
    - **“My code currently won’t scale past 500 PEs”**
    - **“I need a DNS-enlightened closure model”**
  - **Pre/Post-Processing**
    - **“I need a parallel Client-Server viz tool”**
    - **“I am going to generate 5PB of data I need around for 5 years”**
    - **“I need <1 month end-to-end turnaround time to satisfy sponsors”**

Future Requirements

- **Future requirements are best articulated and predicted by understanding the footprint of your application on a given platform**
  - **On-node memory, CPU needs**
  - **Off-node local/global communication**
- **In a normalized sense, this footprint is surprisingly constant.**
  - **Unless/until models and algorithms change**
- **At least this has been my experience**

Questions for You to Consider

- **What challenges do you envision in development (software, algorithms, models) and analysis?**
- **Is your project primarily development, analysis, or both?**
- **Are there any areas of expertise or experience that could preclude your project from succeeding (what keeps you up at night) ?**
- **Is your project of sufficiently high technical risk (Are you counting on "discoveries" to succeed)?**
- **What evolutionary or revolutionary advancements might totally change how you approach problems and the type of problems solved in your field?**
- **How can we best stay abreast of your scientific output? How can we measure your progress? Who are your peer groups?**
- **How can the User Council meet your needs in terms of gathering requirements and communicating regularly?**
- **Why do you need Leadership-Class computers? Increase temporal and/or spatial resolution? Decreased turn-around time? Better models/algorithms? Or?**

NCCS Director of Science: Some R&Rs

- **Delivery and Quality of the Science**
  - Ensure the science is delivered and of the highest (world-leading) quality
  - Develop metrics for & track progress against the quality and delivery
  - Seek and proactively respond to internal and external peer review
- **Align and Develop the Science**
  - Ensure alignment of science thrust areas with DOE thrust areas
  - Stay abreast of & participate in evolutionary & revolutionary science developments
  - Turn new computational science developments into self-sustained programs
  - Help manage the Call for Proposals process
- **Respond to Needs of the Science**
  - Elicit and manage science-based user requirements (User Council)
  - Increase user productivity (code performance, end-to-end analysis time)
- **Communicate the Science**
  - Capture & disseminate science results and software
  - Ensure science output is highly visible
- **Take Chances with the Science**
  - Keep the bar high by taking risks – make sure they are Grand Challenges

This cannot be achieved without your help